

Chapter 4

Perception and Synthesis of Sound-Generating Materials

Bruno L. Giordano and Federico Avanzini

4.1 Introduction

The mechanical properties of sound-generating objects and events in our environment determine lawfully the acoustical structure of the signals they radiate (e.g. Fletcher and Rossing 1991). The ability of listeners to estimate the mechanical properties of everyday non-vocal, non-music sound sources based on acoustical information alone has been the object of empirical research for more than three decades (Vanderveer 1979). Given the lawful specification of the mechanics of the sound source in the acoustical structure, and the adaptive tendency to interpret sensory information in terms of the properties of objects and events in the environment, it is thus not surprising that source-perception abilities are often remarkably accurate (see Lutfi 2007, for a review).

The concept of material is central to the study of source perception from both a theoretical and empirical point of view. The theoretical relevance of this concept originates from the work of Gaver, who outlined a widely influential taxonomy of everyday sound events (Gaver 1993). Accordingly, non-vocal sound sources offer perceptual systems with information about “materials in interaction”, and, at the most general level of the taxonomy, can be classified into three categories depending on the state of matter of the vibrating sound-generating substance: (i) solid sound sources (e.g. clapping); (ii) liquid sound sources (e.g. pouring coffee); (iii) aerodynamic/gaseous sound sources (e.g. wind; explosions). For a variety of reasons (e.g. easiness in manipulation of source mechanics; ecological pervasiveness), the

B.L. Giordano (✉)

Institute of Neuroscience and Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, UK
e-mail: brungio@gmail.com

F. Avanzini

Department of Information Engineering, University of Padova, Via Gradenigo 6/A, 35131 Padova, Italy
e-mail: avanzini@dei.unipd.it

© Springer-Verlag London 2014

M. Di Luca (ed.), *Multisensory Softness*, Springer Series on Touch and Haptic Systems, DOI: 10.1007/978-1-4471-6533-0_4

empirical study of the perception of source mechanics has favoured solid sound sources, and investigated the audition of three different attributes: (i) geometry (e.g. shape of a struck bar; Lakatos et al. 1997); (ii) material (e.g. hardness of a mallet striking a pan; Freed 1990); (iii) properties of the interaction between sound-generating objects (e.g. bouncing vs. breaking of objects; Warren and Verbrugge 1984). The empirical centrality of materials then originates quite simply from the fact that among these three perceptual abilities, material perception has been by far the most studied.

In the first part of this chapter, we will review studies on the perception of material properties from sound. We will initially describe the available empirical evidence on the perception of the state of matter, and then detail the psychophysics literature on the recognition of the material properties of stiff solid objects. Importantly, this last group of studies did not investigate highly compressible solid materials such as soft rubbers, most likely because of the often perceptually negligible acoustical energy they radiate when set into vibration. A subsequent section will detail studies of two classes of deformable materials: fabrics and liquids. Although these investigations addressed the perception of material-independent properties such as texture or liquid amounts, they are summarized here because of the potential interest to future research in the field. The last portion of this part ends with a presentation of studies on the perception and motor-behaviour effects of stiff materials in audio-haptic contexts (see Chap. 2 for visual-haptic contexts).

Studies in ecological perception are the starting point for the development of interactive sound synthesis techniques that are able to render the main perceptual correlates of material properties, based on physical models of the involved mechanical interactions. In the second part of this chapter we will then review recent literature dealing with contact sound synthesis in such fields as sonic interaction design and virtual reality. Special emphasis will be given to softness/hardness correlates in impact sounds, associated to solid object resonances excited through impulsive contact, and rendered using modal synthesis techniques. We will also summarize recent advances in terms of optimization and automation of analysis-synthesis schemes. Two final sections will address less developed literature on the sound synthesis and rendering of deformable objects (notably textiles), aggregate objects (e.g. sand, snow, gravel, and so on), and liquids.

4.2 Perception

4.2.1 *State of Matter*

The first study that gave some indication of the perceptual relevance of the state of matter was carried out by Ballas (1993). Participants in this study were asked to rate a set of 41 environmental sounds that included liquid, aerodynamic (gases), and solid sounds along a variety of scales meant to assess their perceptual representation (e.g. dull vs. sharp timbre) but also aspects of their cognitive processing (e.g. sound

familiarity; similarity of sound to a mental stereotype). Principal component analysis of the rating data outlined three orthogonal judgment dimensions, interpreted as measuring: (i) the identifiability of the sound event (e.g. familiarity scale); (ii) the sound timbre (e.g. dull vs. sharp rating scale); (iii) attributes of the categorical representation (number of similar sounds). A cluster analysis of these principal components revealed one cluster of liquid sounds, which however also included sounds produced in a water context (e.g. boat whistle). The other three clusters included both solid and aerodynamic sources, and grouped together either signals with similar functions (e.g. cluster of signalling sounds, and cluster of door sounds), or highly transient sounds independently of whether they were generated by aerodynamic or solid events (cluster of transient sounds such as a stapler). Overall, this initial study lent some support to the hypothesis that listeners are capable of differentiating between states of matter, although the clustering structure was most likely influenced by a number of factors related to the cognitive processing and higher-order information about the sound signal such as the context where it is generated.

Gygi et al. (2007) investigated the dissimilarity ratings and free sorting of a set of 100 sounds that included living human and non-human sounds (both vocalizations and non-vocalizations), and non-living sounds generated by solid, liquid and aerodynamic sources. In the dissimilarity-ratings experiment participants were presented with all of the possible pairs of stimuli, one pair at a time, and were asked to rate how dissimilar they were. In the free-sorting experiment, participants were presented with all of the stimuli and were asked to create groups of similar sounds. Dissimilarity ratings were analysed with a multidimensional scaling (MDS) algorithm (Borg and Groenen 1997). In general, MDS models the input dissimilarities as the between-stimulus distance within an Euclidean space in which dissimilar stimuli are located further apart. The tendency of stimuli to cluster, i.e., to form tight groups within the MDS space, can thus give an indication of the ability of listeners to differentiate between different-group stimuli (see Ashby 1992, for the relationship between dissimilarity, categorization and discrimination). Overall, both the dissimilarity-ratings and free-sorting data revealed a tendency to group stimuli based on source attributes and, importantly, to differentiate between solid, liquid and aerodynamic events. The tendency to group together sounds generated with substances in the same state was more evident in the free-sorting than in the dissimilarity-ratings data. Indeed, clustering in the dissimilarity-ratings MDS space appeared to be more driven by acoustical attributes that are not always differentiated between diverse sound-generating mechanical systems (e.g. gunshots and footstep sounds were clustered together because they both comprised sequences of transient impact-like sounds). This discrepancy can be explained by the fact that dissimilarity-ratings data are more sensitive to differences in acoustical structure than free sorting data (Giordano et al. 2011).

Further support for the hypothesis of a perceptual relevance of the state of matter was obtained by Houix in an experiment on the free sorting of 60 sounds encountered in a kitchen context (Houix et al. 2012). Vocal sounds were not included in order to eliminate possible distortions in the sorting data independent of the state of matter arising from the likely strong perceptual and attentional salience of the vocal/

non-vocal distinction (Belin et al. 2000; Lewis et al. 2005; Gygi et al. 2007) and from differences between the cognitive processing of living and non-living sounds (Giordano et al. 2010). Consistently with Gygi et al. (2007), participants appeared to group together sounds based on the mechanics of the sound source even when not explicitly required to do so, and created isolated clusters of solid objects, machine and electric device sounds, liquid sounds, and aerodynamic sounds.

Overall, the studies summarized up to this point reveal that the state of matter is likely to structure the cognitive organization of everyday sound sources. These studies, however, do not give evidence concerning the actual ability to recognize the state of matter of a sound-generating substance. A number of sound-generating human-made solid objects have indeed been designed to create “state of matter” illusions (e.g. rainsticks; wind and thunder machines used for centuries in theatres). A recent study by Lemaitre and Heller (2013) addressed this point rigorously. A set of 54 sounds were generated with three different types of interaction for each state of matter (solid: friction, deformation and impact; liquid: splashing, dripping and pouring; gases: whooshes, wind and explosions). Sound duration was gated at different levels. Overall, untrained listeners were able to recognize the state of matter at above-than-chance levels across gating durations and interaction types (75 % correct). Figure 4.1 shows the spectrogram of a set of sound stimuli used in their study.

4.2.2 Perception of Stiff Solid Materials

Except for the recent study by Lemaitre and Heller (2012), investigations into the auditory perception of stiff solid materials were all carried out with isolated impact sounds. In particular, real or simulated impact sounds were generated by the interaction between two objects: the hammer and the sounding object, the former being much more damped than the latter (e.g. when a drum stick strikes a cymbal, the impact sets into vibration the drum stick for a much shorter time than the cymbal). Given the high damping of the former, the sound signal presented to the listeners in these studies contains little or no acoustical energy radiating directly from the hammer, and can be assumed to be the product of acoustical radiation from the sounding object alone. Notably, however, the material properties of the hammer still influence the acoustical structure of the radiated sound signal (Fletcher and Rossing 1991). For example, stiffer hammer materials produce a decrease in the duration of the contact between the hammer and sounding object during the impact, resulting in a more efficient excitation of the high-frequency vibrational modes of the sounding object and, consequently, in an increase in the high-frequency energy of the radiated sound.

Impact sounds can be modeled as the sum of sinusoids whose amplitude decays exponentially starting from the onset of the sound signal. Ignoring perceptually negligible delays, the temporal location of the sound onset essentially corresponds to the time of contact between the hammer and the sounding object. The material properties of the sounding object (elastic coefficients; density), together with its geometry,

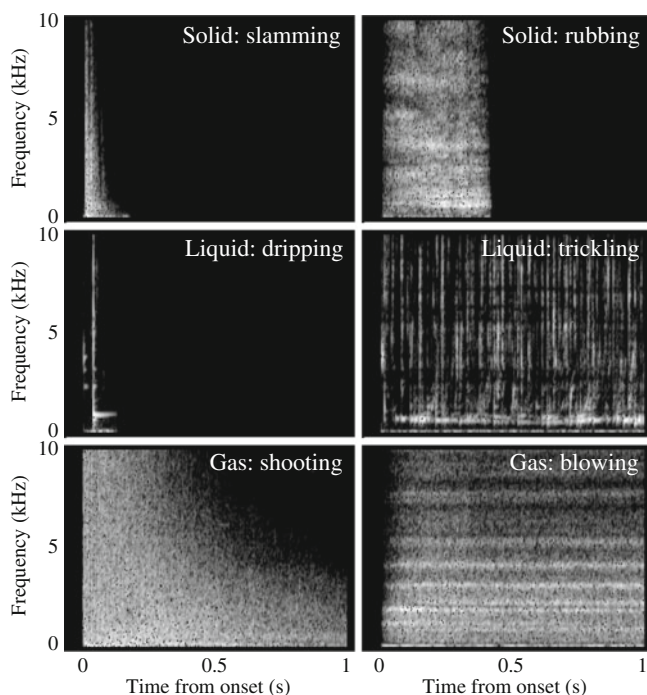


Fig. 4.1 Spectrum of the sound stimuli generated with solid, liquid, and gaseous substances. The *left panels* group impulse-like sounds, generated as a consequence of a temporally localized input of energy in the sound-generating system. The *right panels* show continuous sounds, generated through a temporally distributed input of energy into the sound-generating system. Level from *black* (low) to *white* (high). Data courtesy of Lemaitre and Heller (2013)

determine the frequency of the exponentially decaying spectral components: stiffer and denser materials (and smaller object sizes) produce higher frequency spectral components. The material properties of the hammer, together with the geometry of the hammer and the properties of the hammer-sounding object interaction (e.g. striking force; duration of the hammer/sounding object contact) determine the initial amplitude of the spectral components, and the overall spectral distribution of energy. Stiffer hammers determine higher energy levels in the high-frequency regions. Importantly, the material properties of the sounding object also determine the decay times of the spectral components. Overall, stiffer materials produce spectral components characterised by a slower decay. The velocity of the decay of the spectral components is, however, not constant across all of the spectral components: to a rough approximation, higher-frequency components decay faster than low-frequency components. Wildes and Richards (1988) outlined a simplified yet widely influential model of the relationship between spectral frequency and energy decay. In their model, an increase in spectral frequency produces a linear increase in the decay time of the spectral components. In particular, for stiffer materials the increase in spectral

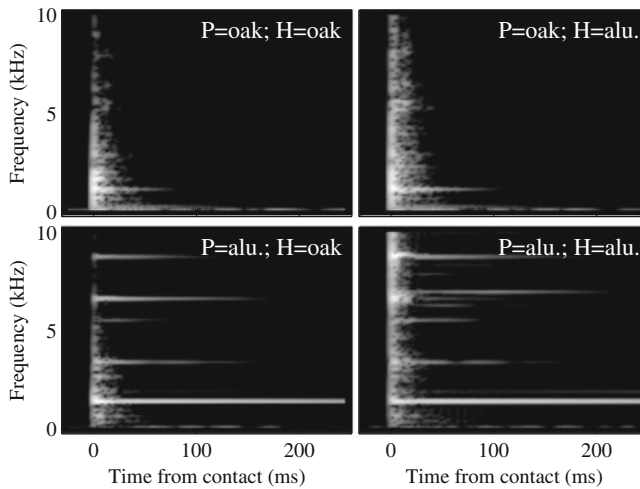


Fig. 4.2 Detail of the spectrotemporal structure of impact sounds generated by striking a 450 cm² square plates (P) made of oak or aluminium with an aluminium or oak hammer (H). Level from low (*black*) to high (*white*). Data from Giordano et al. (2010)

frequency produces a less pronounced decrease in decay time, whereas the contrary is true for more flexible sounding object materials. Within this formulation, the slope of the linear relation between spectral frequencies and decay time is assumed to be a reliable acoustical measure of the stiffness of the sounding object. This variable takes the name of $\tan \phi$, and has been assumed to measure a geometry-invariant acoustical correlate of the sounding-object material. Figure 4.2 displays the time-varying spectrum of the sounds generated by striking an aluminium and oak square plate (area = 450 cm²; thickness = 1 cm) with a semi-spherical aluminium or oak hammer. Note: (i) the longer duration, i.e., lower decay time of the spectral components in the aluminium plate; (ii) the steeper decrease of decay times with frequency for the oak plate; (iii) the higher energy of high-frequency components for the sounds generated with the aluminium hammer.

Two of the earliest studies on the identification of the material of a sounding object were carried out with real sounds, and were primarily concerned with ascertaining identification performance rather than the acoustical factors involved in the identification process. Gaver (1988) struck wood and steel bars of different lengths with a rubber hammer. The vibration of the bars was externally damped with a carpet, on top of which they rested while being struck. Participants were presented with recorded impact sounds, and reached nearly perfect material-identification performance. Kunkler-Peck and Turvey (2000) investigated the ability to identify simultaneously the material and shape of a freely-vibrating plate (sounding object) struck with a steel pendulum (hammer). All of the plates had the same area, but differed in both shape (circle, triangle, rectangle) and material (steel, wood, Plexiglas). Sounds were generated live by the experimenter, while being occluded from the view of

the experiment participant. The identification of material was virtually perfect (only one misidentification was recorded) and was more accurate than the identification of shape which, however, was significantly better than chance.

Another set of early studies on the perception of the material properties of sounding object focused on synthetic sounds (Avanzini and Rocchesso 2001; Lutfi and Oh 1997; Klatzky et al. 2000). Stimuli in these studies were synthesized as the sum of exponentially decaying sinusoids. In the studies by Avanzini and Rocchesso (2001) and by Klatzky et al. (2000), material properties were controlled by manipulating $\tan \phi$. Participants in these studies were either asked to identify sounding-object materials (Avanzini and Rocchesso 2001; Klatzky et al. 2000), or to discriminate the change in material between two subsequently presented sounds (Lutfi and Oh 1997). Based on the work by Wildes and Richards (1988), participants' responses were hypothesized to focus on the acoustical information contained in the decay of the spectral components, because it would specify materials more reliably than their frequency. In practice, however, participants' responses were influenced by both decay and frequency information. In the study by Lutfi and Oh (1997) the reliance of the responses of some participants on frequency information was so strong that it effectively overshadowed the perceptual effect of decay.

The empirical observation that the auditory perception of materials can be strongly influenced by sound frequency was problematic because it contradicted the theoretical framework that dominated the field, the ecological approach to perception (Gibson 1966, 1979; Michaels and Carello 1981). Accordingly, it was thought that source perception would rely on the detection of invariants, i.e., parameters of the acoustical structure that specify reliably the source property under judgment independently of variations in non-target source properties (e.g. an invariant for sounding object material specifies this property reliably and independently of changes in non-material properties such as size). Since acoustical parameters such as $\tan \phi$ were thought to constitute invariant information about materials, it was surprising that the identification of material was also influenced by sound frequency, a variable that is influenced by both the material and geometry of the sounding object (Fletcher and Rossing 1991). Carello et al. (2003) argued that the focus of listeners on frequency information was an artefactual product of the synthetic nature of the sound signals which provided participants with impoverished material-related information. Giordano and McAdams (2006; Giordano 2003) addressed this issue within a material-identification study conducted with real sounds recorded by striking wood, plastic, metal and glass plates of different sizes. Importantly, an analysis of the acoustical structure revealed that sounds were rich in information that differentiated between materials independently of variations in size. Among them was a psychoacoustically plausible derivation of the $\tan \phi$ coefficient. When asked to identify materials, however, participants did not appear to fully exploit this invariant acoustical information. Indeed, when it came to differentiating between gross material categories, i.e., wood and plastic on the one hand, and metal and glass on the other, identification performance was perfect and could be accounted for by a focus on various acoustical features among which $\tan \phi$. However, identification performance within each of these two categories was virtually at chance level because participants differentiated

between metal and glass or between wood and plastic by relying exclusively on the size of the plate or, from the acoustical point of view, on the lowest spectral component. Identification confusions between wood and plastic were also observed by Tucker and Brown (2003) in a study on real sounds, and were instead inconsistent with the results by Kunkler-Peck and Turvey (2000) obtained with sounds generated live. It is unclear whether participants in the latter study were exposed to additional material-related information through eventual uncontrolled sounds generated while the experimenter hung the plates on the apparatus used to strike them.

The effect of frequency information on identification responses in the study by Giordano and McAdams (2006) was consistent with previous studies of synthetic sounds, and, together with the observation of the availability of acoustical information for the size-independent identification of material, disconfirmed the impoverished information hypothesis for the perceptual focus on sound frequency. Various hypotheses can be advanced to explain the influence of object size and sound frequency on the identification of the material of an object. It might be for example argued that participants in the study by Giordano and McAdams erroneously interpreted all of the available acoustical information in terms of object material because they were not informed of the variation in the size of the plates. This hypothesis would, however, not explain why participants in this study identified small glass and metal plates, which generated higher-frequency sounds, as being made of glass because metal is denser and stiffer than glass (Waterman and Ashby 1997) and should thus generate higher-frequency sounds. Another possible interpretation for the reliance of material identification on sound frequency is more subtle. Studies on the perception of musical timbre show that the influence of pitch on judgments of the dissimilarity of musical timbres grows with the range of variation of pitch within the experimental set and, in particular, becomes relevant at the expense of a focus on non-pitch acoustical information when participants are exposed to pitch variations larger than one octave (Handel and Erickson 2001; Marozeau et al. 2003; Steele and Williams 2006). Accordingly, the reliance of material identification on frequency observed in previous studies was determined by the comparatively large variation of this pitch-related acoustical variable (e.g. six octaves in Giordano and McAdams 2006). Consistently with this interpretation, a subsequent material-identification experiment by McAdams et al. (2010), carried out with a sound set that included a much smaller frequency variation (less than half an octave) revealed no effect of this variable on identification responses (see McAdams et al. 2004; Giordano 2005; McAdams et al. 2010 for dissimilarity-ratings studies of impacted sound sources). A final plausible explanation of the effect of frequency on identification responses is that it is the product of the internalization of a statistical regularity in the acoustical environment. Accordingly, listeners identified the high-pitched sound of a small metal plate as being made of glass because of the small size of everyday freely vibrating glass objects (e.g. clinking glasses).

Overall, studies on the identification of the material of impacted sounding objects reveal a nearly perfect ability to differentiate between gross categories of materials (wood or plastic vs. metals or glass), and a number of frequency-dependent biases in the identification of materials within these categories. From the acoustical point

of view, the ability to differentiate between these gross categories appears to be well explained by the perceptual processing of an acoustical measure of the damping of vibrations, $\tan \phi$ (Giordano and McAdams 2006; McAdams et al. 2010; Avanzini and Rocchesso 2001; Klatzky et al. 2000). Notably, however, this acoustical variable might not account well for the differentiation of gross material categories in non-impact interaction types (Lemaitre and Heller 2012). From the mechanical point of view, the ability to differentiate between gross material categories appears to be robust to variations in the geometry of the struck object (see also Ren et al. 2013), and vulnerable only to the external damping of the vibration of the sounding object, either when it is submerged underwater (cf., identification of material in sounds recorded under water; Tucker and Brown 2003), or when it is attached to a dampening soft-plastic surface (Giordano 2003).

The studies reviewed up to this point investigated the identification or discrimination of material categories, but did not assess the ability to estimate quantitative material attributes such as their hardness/softness. Giordano et al. (2010) assessed the ability of participants to discriminate and rate the hardness of sounding objects in the presence or absence of training. Sounds were generated by striking variable size plates of different materials with hammers of different materials. In an initial experiment, listeners received correctness feedback when asked to discriminate the hardness of sounding objects. Within a limited number of blocks trials they were able to quickly learn to discriminate the hardness of sounding objects independently of variations in its size, and independently of variations in the hardness of the hammer. The training received in the discrimination experiment generalized to a second hardness-rating experiment where they did not receive correctness feedback. Also in this experiment they were able to accurately rate the hardness of sounding objects independently of variations in their size and of variations in the material of the hammer. Importantly, another group of participants who did not receive prior discrimination training appeared to estimate the hardness of sounding objects by focusing on the target mechanical properties, but did not ignore the hardness of the hammer and the size of the sounding object which still influenced their rating responses although to a lesser extent. Similar effects were obtained with another group of untrained listeners who rated the hardness of sounding objects when presented with synthetic impact sounds. Again, their estimates were most strongly influenced by the synthetic parameter modeling the hardness of the sounding object, but they were influenced, to a lesser extent, by the frequency of the sounds, and by the acoustical parameters that in real sounds were most strongly influenced by the impact properties. Overall, the study by Giordano et al. (2010) confirms the tendency of untrained listeners to estimate the material properties of the sounding object by considering also non-material parameters such as the size of the sounding objects. The fact that trained listeners are able to estimate the hardness of sounding objects independently of their size confirms the presence of perceptually available yet not completely exploited information for accurate material perception.

Only three studies investigated the ability to perceive the material properties of the hammer (Freed 1990; Lutfi and Liu 2007; Giordano et al. 2010). In the study by Freed (1990), participants were presented with the sound of variable-sized metallic

pans struck with mallets of different hardness and were asked to estimate the hardness of the hammer. Hardness estimates appeared to be accurate, and independent of the size of the sounding object. Lutfi and Liu (2007) investigated the weighting of the amplitude information of the spectral components of synthetic impact sounds within a hammer-hardness discrimination task, and assessed the extent to which the weighting strategies are reliable across different days. Although different individuals were characterized by largely diverse patterns of information weighting, the weighting strategies of each single individual were highly replicable.

The study by Freed lent support to the hypothesis that participants are extremely accurate at perceiving the material properties of hammers. Combined with the results concerning the often imperfect identification of the material of sounding objects, it would thus appear, paradoxically, that the auditory system is better equipped at detecting the properties of an object that only indirectly structures the acoustical signal through its effects on the sounding object vibration—the hammer—rather than at detecting the material properties of the sound-radiating sounding object. However, it should be noted that in Freed (1990) the material of the hammer varied while that of the sounding object was kept constant. As such, this study offered only a rather limited test of the extent to which the auditory perception of hammer materials is truly invariant. In a follow-up study carried out with both real and synthetic sounds, Giordano et al. (2010) asked participants to estimate the hardness of hammers that struck plates of different size and material. Across multiple conditions, participants were able to estimate accurately the hardness of the hammer only when receiving trial-by-trial feedback on discrimination performance. In the absence of such a feedback, they appeared instead to estimate hammer hardness based on properties of the hammer-sounding object impact, such as the duration of their contact. Importantly, impact properties are influenced by both the material of the hammer and of the sounding object, i.e., perception of the hammer material relied on a less-than optimal mechanical variable.

Overall, studies on the perception of solid materials have revealed that in the absence of explicit training or feedback on the correctness of their responses, the perceptual abilities that listeners bring into the experimental context are often less than perfect, and are influenced by non-target mechanical properties of the sound source. These results make it rather unlikely that naive listeners rely on acoustical invariants that specify accurately a target source property. In line with this interpretation, it has been frequently observed that listeners perceive sound source properties by relying on multiple attributes of the acoustical signal:

- the perception of the geometrical properties of the sounding object is influenced by the frequency of the spectral components (Lakatos et al. 1997; Lutfi 2001; Houix 2003), by the properties of the sound decay (Lutfi 2001; Houix 2003), and by the distribution of energy across the spectrum, as measured by the spectral centroid, the amplitude-weighted average of the spectral frequencies (Lakatos et al. 1997);
- the perception of the material properties of the sounding object is influenced by the properties of the sound decay (Lutfi and Oh 1997; Klatzky et al. 2000; Avanzini and Rocchesso 2001; Giordano and McAdams 2006; McAdams et al. 2010), but also

by sound frequency (Klatzky et al. 2000; Avanzini and Rocchesso 2001; Giordano and McAdams 2006);

- the perception of the material and mass of the hammer is influenced by both loudness and spectral centroid (Freed 1990; Grassi 2005).

The empirical observation of the reliance of perceptual judgment on multiple features of the sound signal opens up the question of how listeners establish their perceptual weight, i.e., the strength of their influence on the perceptual estimation of the source property. Giordano et al. (2010) investigated the extent to which the perceptual weight of an acoustical feature could be accounted for by two different principles. Firstly, the accuracy of the acoustical information, i.e., the extent to which an acoustical feature specifies accurately the mechanical property. Secondly, the “exploitability” of the acoustical information, i.e., the extent to which the perceptual system can use the available acoustical information given limitations in discrimination, learning and memory.

The information-accuracy principle is a quantitative extension of the invariant-information hypothesis originating within the ecological approach to perception. Accordingly, an ideal observer that carries out a source-perception task (e.g. ratings of the hardness of sounding objects) can achieve a different performance level when focusing on different acoustical features. A given acoustical feature is thus characterized by a specific task-dependent accuracy score, i.e., it affords a given performance level that ranges from chance level to perfect (e.g. zero to perfect correlation between hardness ratings and actual hardness levels). Giordano et al. (2010) measured such task-dependent information-accuracy scores by analysing a large database of impacted sounds, and hypothesized an increase of perceptual weight with an increase in their value. The information-exploitability principle states instead that, independently of the task at hand, the perceptual response will be more strongly influenced by acoustical features that, in general terms, are processed more efficiently by the observer. For example, acoustical features that are better discriminated will have a stronger influence on the estimation of a given source property (e.g. Ernst and Banks 2002). Similarly, perceptual weights will be higher for acoustical features that observers learn more quickly to associate with a given source property, or whose association with a source property is stored in memory more stably. In Giordano et al. (2010), information exploitability was measured by the ability of listeners to retain and generalize the perceptual focus on a given acoustical features from a condition where they received trial-by-trial feedback on response correctness to a subsequent condition where such feedback was not available.

In an initial discrimination experiment, Giordano et al. (2010) observed that listeners learn quickly to discriminate the hardness of sounding objects, whereas they require a longer training to reach the same target performance level when discriminating the hardness of hammers. When they received trial-by-trial performance feedback participants focused on the most accurate acoustical features for the discrimination of the hardness of both objects. However, in the absence of such a feedback the same participants in a second hardness-rating experiment were able to retain the focus on the most accurate information only when it came to estimating the hardness of sounding

objects, but not of hammers. As such, the perceptual focus on accurate information appeared to be strongly limited by the ability of participants to learn and retain in memory perceptual criteria acquired while carrying out the initial discrimination task. Overall, this study thus shows that the weighting of acoustical information for the perception of sound sources is not always dominated by the accuracy of acoustical information, but is also determined by task-independent limitations in the processing abilities of the perceptual system. This view is consistent with the observation by Lutfi and Stoelinga (2010) that performance in the perception of the properties of a struck bar can be accounted for by the ability of listeners to discriminate the features of the sound signal.

4.2.3 Comparison of Material and Interaction Perception

Everyday non-vocal sound sources can be differentiated based on both the sound-generating materials and the type of interaction that sets them into vibration: solid materials can be set into vibration by plastic deformations (e.g. crumpling paper), impacts, scraping and rolling; liquid interactions include dripping, pouring, and splashing; interactions for aerodynamic/gaseous sound sources include explosions, gusts, and wind-like turbulence (Gaver 1993). Given the centrality of the construct of interaction to the organization of everyday sound sources, and the strong effects they have on the structure of the acoustical signal (see Figs. 4.1 and 4.3), it is thus natural to ask which, among materials and interactions, are central to the cognitive

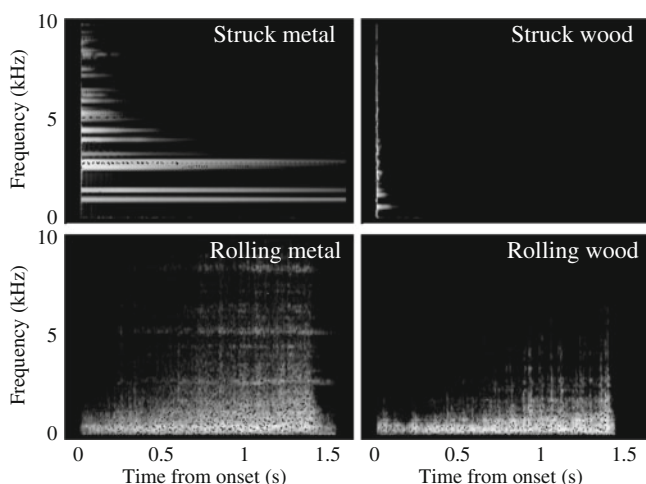


Fig. 4.3 Spectrum of the sound stimuli generated by hitting or rolling the same metal (*left panels*) and wood (*right panels*) cylinders. Level from low (*black*) to high (*white*). Data courtesy of Lemaitre and Heller (2012)

organization of everyday non-vocal sounds, and which our auditory system can process better. Lemaitre and Heller addressed these questions in two recent studies (Lemaitre and Heller 2012, 2013).

In Lemaitre and Heller (2013), participants were presented with sounds generated by setting into vibration various substances (solids, gases and liquids) with different types of actions. They carried out a label-verification experiment, i.e., they had to decide whether a given label was an appropriate description of the sound-generating event. Labels could belong to one of five categories: (i) state of matter; (ii) interaction type (e.g. friction for scraping and squeaking or deformation for tearing and crumpling); (iii) specific interaction (e.g. scraping); (iv) manner of action (e.g. scraping rapidly); (v) object of the action (e.g. scraping a board). Labels belonging to the specific-interaction category were verified more quickly and accurately than labels from any other category, suggesting a central role of this description level in the cognitive organization of the sound stimuli. The same conclusion was supported by the results of a second priming experiment where sounds primed a lexical decision task carried out on subsequently presented labels belonging to either the interaction type or specific interaction category. Responses were faster and more accurate for specific-interaction labels.

In Lemaitre and Heller (2012), sounds were generated by applying different interaction types (rolling, scraping, hitting and bouncing) to hollow cylinders made of four different materials (plastic, wood, metal and glass). In an initial experiment, participants rated how well the sound conveyed either a specific interaction type or a specific material. Performance measures derived from the ratings data were significantly better when participants judged the interaction type. In a second experiment, participants were asked whether a target sound had been generated with a given interaction type or material described by a label presented before the sound onset. Reaction times were faster for the identification of the interaction type.

Overall, the studies by Lemaitre and Heller support the hypothesis that the manner in which substances and objects are set into vibration play a more central role in the cognitive organization of non-vocal everyday sounds. Further, interaction types appear to be associated with acoustical fingerprints that are processed more quickly by the auditory system than those that characterize materials and states of matter.

4.2.4 Perception of Deformable Materials

Sounds produced by fabrics are the result of complex interactions that include sliding friction but also crumpling sounds due to buckling of the fabric on itself. The frictional component of these sounds is largely influenced by the texture of the fabric, a parameter mostly related to the fine-grained geometry of the object rather than to geometry-independent material properties. The reader is referred to the work by Lederman and co-workers on the audio-haptic perception of texture (Lederman 1979; Lederman and Klatzky 2004).

In the scientific literature there appears to be a lack of studies on the auditory perception of material-related properties of fabrics and textiles. On the other hand, the definition of objectively measurable properties of fabrics (the term Fabric Objective Measurement—FOM—is often used) is a central issue in the field of textile and apparel industry. For this reason, the perception of fabrics and textiles has been studied in the context of the evaluation of fabric “hand”, quality, and related performance attributes. Judgments of hand feel properties of fabric are traditionally given by panels of experts, and sets of hand feel properties and corresponding scales have been defined (Civille and Dus 1990), most of which are related to the tactile sense. Among them, surface softness (on a subjective scale *soft* → *hard*) and, interestingly, two auditory properties: “noise intensity” (*soft* → *loud*), and “noise pitch” (*low/bass* → *high/sharp*). The Kawabata evaluation system (KES) (Kawabata 1980) has been developed under the assumption that the main characteristics of fabric responsible for hand feel depend on many physical properties, including dimensional changes at small forces (tensile, shear, compression, bending), surface properties (roughness and friction), and surface thermal insulation. KES testing instruments estimate various mechanical properties of fabric, which are then related to sensory signals acquired through hand-contact.

Within this research area, the role of auditory information in the subjective evaluation of fabrics has been assessed in a series of studies on the effects of sound on hand-feel properties (Cho et al. 2001, 2005). In Cho et al. (2001), a set of psychophysical experiments was presented concerning the characteristics of frictional sounds produced by interaction with fabrics, and their effect on the perceived quality of fabrics. Among various fabric-sound features, the sound “level range” ΔL was found to exhibit a positive correlation with perceived softness and pleasantness. In Cho et al. (2005), the reaction of observers to frictional sounds of warp-knitted fabrics was measured in terms of physiological responses (EEG, respiration rate, skin conductance level, etc.). In this study, the psychoacoustical measures of roughness and fluctuation strength were strongly correlated with the perceived pleasantness of fabric sounds.

Although liquid sound sources have been investigated in a number of studies, few of them assessed the ability of listeners to perceive a mechanical property of the sound source. The most active area of research on liquid sounds is indeed of an applied nature, and aims for example to assess how they can be used to mask road-traffic noise (e.g. De Coensel et al. 2011; Jeon et al. 2012). Overall, this research field shows promising potentials for the improvement of the perceptual quality of urban soundscapes. Consistently, a semantic-differential study of environmental sounds carried out by Kidd and Watson (2003), revealed that across a large set of environmental sounds, liquid sounds are among the least harsh (e.g. splashing vs. breaking light bulb sound) and the most appealing (e.g. waterfall vs. scraping wood sounds).

Although not directly related to the auditory perception of the mechanics of the sound source, the study carried out by Geffen et al. (2011) gives interesting indications on how liquid sounds might be encoded in the auditory system. It is known that the acoustical structure of natural sounds exhibit scale-invariant or self-similarity

traits, as measured by the $1/f$ distribution of the power spectrum of amplitude fluctuations across different frequencies f (Voss and Clarke 1975; Attias and Schreiner 1997). Confirming these earlier studies, Geffen et al. (2011) observed that natural water sounds exhibit the same scale-invariant $1/f$ spectrum. More importantly, they observed that natural water sounds whose scale was modified simply by altering their playback speed were still perceived as natural and water like and that synthetic stimuli given by the overlap of temporally and spectrally distributed chirps were still perceived as natural and water like only when they exhibited the same scale-invariant structure and a $1/f$ spectrum.

Four studies examined the perceptual estimation of the mechanical properties of liquid sound sources. Jansson (Jansson 1993; Jansson et al. 2006), carried out a series of experiments to assess the estimation of the amount of liquids inside a shaken vessel. In Jansson (1993), participants estimated the amount of liquid in a shaken opaque container in haptic, auditory or visual conditions. Although auditory and visual estimates of the amount of liquid scaled to the actual amount in all conditions, the most accurate estimates were given when haptic information was available. Interestingly, accuracy in all of these conditions improved dramatically after participants were exposed to a prior multisensory condition where information from all of the three modalities was available, revealing a rapid calibration of the processing of information within each of the modalities. In Jansson et al. (2006), participants estimated the amount of liquid or of a solid substance held within a shaken vessel in various haptic conditions, each characterized by a different constraint on the exploratory movement that participants could execute, and in a trimodal visual-auditory-haptic condition. Consistent with the previous study, accuracy was higher in the trimodal condition and in the haptic condition where participants were allowed to shake the vessel, as opposed to only lift it. Cabe and Pittenger (2000), investigated the ability to perceive the filling level in water container. An initial auditory-only experiment revealed that listeners can accurately differentiate between liquid-pouring events where the overall level of water decreases, increases or remains constant. In a second experiment, participants were asked to fill a container up to a specified level. Accuracy was higher when participants had access to haptic, visual and auditory information as compared to when they had access to auditory information only. In a final experiment, inspired by studies on the visual estimation of time to contact, listeners were found able to predict accurately the time it would have taken for a vessel to fill completely after having heard the sound generated by filling it at various below-brim levels. Finally, Velasco et al. (2014) investigated the perception of the temperature of poured liquids. In an initial experiment, sounds were generated by pouring cold or hot water in one of four different containers (glass, plastic, ceramic and paper). Participants were very accurate at identifying whether the poured water was cold or hot for all containers. Good recognition abilities were confirmed in a second experiment that measured the implicit association between cold and hot sounds, and the “cold drink” and “hot drink” verbal labels. In a final experiment, participants rated the temperature of one cold and one hot liquid pouring sound, and of a manipulated version of the cold sound that increased high-frequency energy, and of the hot sound that increased low-frequency energy. Temperature ratings increased from the manipulated cold to the

original cold to the original hot to the manipulated hot sound, suggesting an increase in perceived liquid temperature with a decrease in the low-frequency energy of the sound signal.

4.2.5 Audio-haptic Perception of Materials

The studies reviewed up to this point on the perception of stiff impacted materials reveal that geometry-independent material information is highly relevant to the auditory perception of objects and events. In contrast, it is a largely open question whether auditory material-related information is perceptually relevant when presented in a multisensory context because relatively few studies addressed this question. Chapter 2 by Klatzky and Wu in this volume reviews studies on the integration of audio-visual information about materials. Here, we review studies on the audio-haptic processing of material information. They can be divided in two groups according to the exploratory gesture used by observers to generate the sound signals: hitting and walking.

The audio-haptic perception of struck materials was investigated in three studies (DiFranco et al. 1997; Avanzini and Crosato 2006b; Giordano et al. 2010). DiFranco et al. (1997) investigated the ranking of the stiffness of simulated haptic surface when presented along with recorded sounds generated by striking materials of different stiffness. Stiffness rankings increased with the actual stiffness of the auditory or haptic objects when they were presented along with a constant haptic stiffness or when presented alone, respectively. Two groups of observers participated in the main experiment. They had either taken part in a previous experiment on the ranking of haptic stiffness (expert observers) or not (naive observers). Naive observers appeared to be more strongly influenced by auditory stiffness than expert observers (accuracy in ranking of haptic stiffness across sound-stiffness levels = 44 and 73 % correct, respectively). Notably, however, also the expert observers appeared to take into account auditory stiffness to some extent because their performance in the ranking of haptic stiffness appeared to decrease relative to what was observed when they were not exposed to simultaneous sound stimuli (83 % correct in the haptic-only condition). Avanzini and Crosato (2006b) investigated the perceptual effectiveness of a haptic-synthesis engine coupled with a real-time engine for the synthesis of impact sounds. The sound-synthesis engine in this study allowed the manipulation of the force-stiffness coefficient, a mechanical parameter influenced by the stiffness of both the hammer and sounding object that influences primarily the perceived hardness of the hammer but also of the sounding object (Giordano et al. 2010). Variable-stiffness sounds were presented along with a simulated haptic surface of constant stiffness. Consistently with what observed by DiFranco et al. (1997) with real sounds, the ratings for the stiffness of these audio-haptic events increased with the auditory stiffness.

The same model investigated in Avanzini and Crosato (2006b) was adopted by Giordano et al. (2010) in a study on the effects of audio-haptic stiffness on the motor

control of striking velocity. During the initial phase of each trial, participants received continuous feedback on whether their striking velocity was within a target range. After they reached a given performance criterion, feedback was removed during a subsequent adaptation phase. During a final change phase, the audio or haptic stiffness was modified from the baseline value. Three groups of individuals (non-musicians, non-percussionist musicians and percussionists) participated in four experimental conditions: (i) auditory only; (ii) haptic only; (iii) audio-haptic congruent, where the audio and haptic stiffness was changed in the same direction (e.g. increase in both); (iv) audio-haptic incongruent. Overall, an increase in audio-haptic stiffness led to a decrease in the striking velocity during the change phase of each trial. Notably, however, whereas both non-musicians and musicians decreased striking velocity for an increase in haptic stiffness, percussionists had the exact opposite weighting of this variable, and struck stiffer haptic objects faster. In the audio-haptic condition, the control of striking velocity appeared to be dominated by changes in haptic stiffness. Also, congruency modulated the motor effect of audio stiffness, which had a significant effect only during the audio-haptic congruent condition, whereas it did not modulate the motor effects of haptic stiffness.

Three recent studies investigated the perceptual and motor effects of the properties of walked-upon materials (Giordano et al. 2012; Turchet et al. 2014; Turchet and Serafin 2014). Giordano et al. (2012) carried out an experiment on the non-visual identification of real walking grounds in audio-haptic, haptic, kinaesthetic and auditory conditions. Eight ground materials were investigated: four solid materials (vinyl, wood, ceramic and marble) and four aggregate materials (gravels of four different sizes; see Fig. 4.4 for example waveforms and spectrograms of the sound stimuli in these studies). Three of the experimental conditions were interactive, i.e., participants carried out the identification task after walking blindfolded on the ground material. In the audio-haptic condition, they had access to all of the available non-visual information. In the haptic and kinaesthetic conditions, auditory information was suppressed by means of a masking noise reproduced over wireless headphones. In the kinaesthetic condition, tactile information about ground materials was suppressed by reproducing a tactile masker through a recoil-type actuator installed in an outer sole strapped under the shoe. In the auditory condition participants did not walk on the ground materials, and heard the walking sounds they had generated during the audio-haptic condition. Given the large differences between the vibratory signals generated while walking on solid vs. aggregate materials (see Fig. 4.4), it is not surprising that in all sensory conditions participants almost perfectly discriminated between these two classes of walking grounds. Within each of these categories, identification performance varied across experimental conditions, and was maximized when participants had access to tactile information in the haptic and audio-haptic conditions. In particular, tactile information appeared to be critical for the identification of solid materials because when it was suppressed during the kinaesthetic condition identification was at chance level. More interesting were the results of the analysis of the dominance of the different sensory modalities during the audio-haptic condition. For solid materials, some evidence emerged concerning the dominance of haptic information, i.e., the sensory modality that allowed the

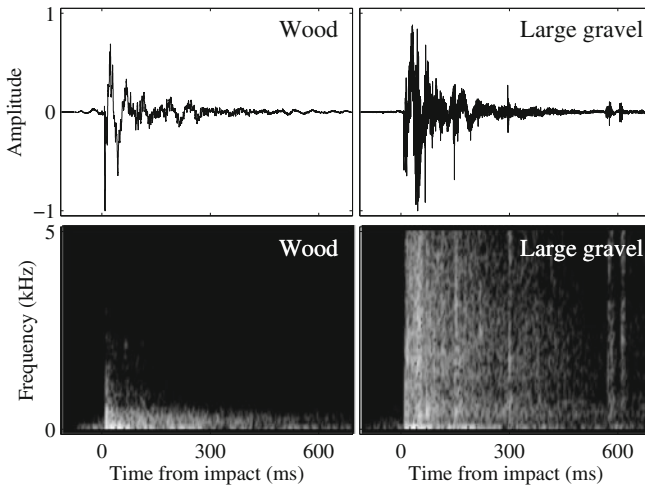


Fig. 4.4 Waveform (*top panels*) and spectrum (*bottom panels*) of a footstep sound generated while walking on wood or on large gravel, an aggregate material. Level from low (*black*) to high (*white*). Time from impact estimated approximately based on waveform. Amplitude scaled to maximum for display purposes. Data from Giordano et al. (2012)

best identification performance appeared to dominate the audio-haptic identification of solid grounds. Surprisingly, the identification of aggregate materials during the audio-haptic conditions appeared instead to be dominated by kinaesthetic information, i.e., participants focused on the worst performing sensory modality. This result was interpreted as revealing a bias in the weighting of modality-specific information: when walking on unstable grounds, such as a gravel, participants were likely not concerned with discriminating accurately the size of the gravel, but were instead concerned with keeping a stable posture by focusing on the sensory modality that would have most promptly signalled a potentially unstable posture, kinaesthesia.

Turchet and Serafin (2014) carried out a study on the congruence of simulated auditory and haptic walking ground. Participants rated the congruence of auditory and haptic materials presented simultaneously either in an active-walking condition or in a passive conditions during which they experienced the audio-haptic display while sitting on a chair. Audio-haptic congruence appeared to be maximized when both materials belonged to either the solids or aggregate category, thus confirming the perceptual relevance of the same distinction observed by Giordano et al. with real grounds. Turchet et al. (2014) finally assessed the extent to which auditory information about ground materials influences the kinematics of locomotion. Participants were instructed to walk using their normal pace on an asphalt ground. Importantly, participants could hear either the sounds they generated while walking on asphalt, or synthetic walking sounds generated in real time with a model meant to simulate wood, gravel, and snow-covered grounds. Auditory information about materials appeared to affect a number of variables related to the pace. For example, pace appeared to be

slower on aggregate than on solid materials (synthetic snow and gravel vs. synthetic wood and real asphalt ground).

Overall, auditory material-related information influences both the perception of materials in audio-haptic contexts and the kinematics of sound-generating movements (see also Castiello et al. (2010), for effects of sound information on the kinematics of grasping). The available experimental evidence appears however to show that within an audio-haptic context auditory materials have weaker effects than haptic materials on both perceptual judgment and motor behaviour.

4.3 Synthesis

In light of the discussion developed in the first part of this chapter, it can be stated that appropriate sound synthesis techniques for the rendering of auditory correlates of material-related sound source properties must possess two main qualities: (i) they have to provide access to sound control parameters that can be related to ecological properties of the simulated sound-generating phenomena, and (ii) they have to be usable in real-time interactive settings, responding naturally to user actions.

Recent literature in virtual reality and sonic interaction design (Rocchesso 2011; Franinović and Serafin 2013) has explored the use of “physically based sound modeling” techniques to develop interactive sound synthesis schemes. This term refers to a set of synthesis algorithms that are based on a description of the physical phenomena involved in sound generation, whereas earlier techniques are based on a description of the sound signal (e.g. in terms of its waveform or its spectrum) and make no assumptions on the sound generation mechanisms.

Since physically based models generate sound from computational structures that respond to physical input parameters, they automatically incorporate complex responsive acoustic behaviours. A second advantage is interactivity and ease in associating motion to sound control. As an example, the parameters needed to characterize impact sounds (e.g. relative normal velocity), are computed in a VR physical simulation engine and can be directly mapped into control parameters, producing a natural response of the auditory feedback to user gestures and actions. Finally, physically based sound models can in principle allow the creation of dynamic virtual environments in which sound-rendering attributes are incorporated into data structures that provide multimodal encoding of object properties (shape, material, elasticity, texture, mass, etc.). In this way, a unified description of the physical properties of an object can be used to control the visual, haptic, and audio rendering (Avanzini and Crosato 2006b; Sreng et al. 2007).

4.3.1 Modal Sound Synthesis

Various physically based modeling techniques exist in the literature, particularly for musical instruments (see e.g. Smith 2004; Välimäki et al. 2006, for extensive

reviews). A full physical simulation entails the numerical resolution of a set of partial differential equations describing mechanical and/or fluid-dynamic oscillations, as well as sound radiation in space. Finite-difference and finite-element models have been used to simulate musical instruments using this approach (Bilbao 2009). In particular, Chaigne and co-workers have developed accurate models for sound-producing mechanisms involving impacted resonators (strings, bars, plates) (Doutaut et al. 1998; Lambourg et al. 2001). A similar approach has been proposed also in the context of non-musical sound synthesis (O'Brien et al. 2002): finite-element simulations are employed for the generation of both animated video and audio. Complex audio-visual scenes can be simulated, but heavy computational loads still prevent real-time rendering and the use of these methods in interactive applications.

A more efficient technique is modal sound synthesis (Adrien 1991). Starting with the studies by Van den Doel and co-workers (van den Doel and Pai 1998; van den Doel et al. 2001), this has become the most used approach for the simulation of non-musical sounds produced by mechanical contact of solid objects. Consider a resonating object described as a network of N masses connected with linear springs:

$$M\ddot{y}(t) + Ky(t) = f_{\text{ext}}(t), \quad (4.1)$$

where y is a vector containing the displacements of the N points of the network, the mass matrix M is typically diagonal, while the stiffness matrix K is in general not diagonal because the points are coupled through springs. The homogeneous equation ($f_{\text{ext}} \equiv 0$) has in general N modal solutions of the form $y(t) = s \cdot \sin(\omega t + \phi)$, where the vector s of the *modal shapes* is an eigenvector of the matrix $M^{-1}K$ with associated eigenvalue ω^2 . The eigenvectors are orthogonal with respect to the mass and the stiffness matrix, and their associated matrix $S = [s_1 | s_2 | \dots | s_N]$ defines a change of spatial coordinates that transforms system (4.1) into a set of N uncoupled oscillators:

$$M_q \ddot{q} + K_q q = S^T f_{\text{ext}}(t), \quad \text{with} \quad M_q = S^T M S, \quad K_q = S^T K S. \quad (4.2)$$

Due to orthogonality, the matrices $M_q = \text{diag}\{m_n\}_{n=1}^N$ and $K_q = \text{diag}\{k_n\}_{n=1}^N$ are diagonal. Therefore the *modal displacements* $\{q_n\}_{n=1}^N$ obey a second-order linear oscillator equation with frequencies $\omega_n^2 = k_n/m_n$, where m_n and k_n represent the modal masses and stiffnesses, and where the transposed matrix S^T defines how a driving force f_{ext} acts on the modes. The oscillation $y_l(t)$ at the l th spatial point is the sum of the modal oscillations weighed by the modal shapes: $y_l(t) = \sum_{n=1}^N s_{n,l} q_n(t)$.

Equivalently, modal decomposition can be obtained from the partial differential equation that describes a distributed object, in which the displacement $y(x, t)$ is a continuous function of space and time. In this case, a normal mode is a factorized solution $y(x, t) = s(x)q(t)$. As an example, for a string with length L and fixed ends, the D'Alembert equation with fixed boundary conditions admits the factorized solutions $y_n(x, t) = s_n(x)q_n(t) = \sqrt{2/L} \sin(\omega_n t + \phi_n) \sin(k_n x)$, with $k_n = \frac{n\pi}{L}$ and $\omega_n = ck_n$ (c is the wave speed). If a force density $f_{\text{ext}}(x, t)$ is acting on the string, the equation is

$$\mu \frac{\partial^2 y}{\partial t^2}(x, t) - T \frac{\partial^2 y}{\partial x^2}(x, t) = f_{\text{ext}}(x, t), \quad (4.3)$$

where T, μ are the string tension and density, respectively. Substituting the factorized solutions $y_n(x, t)$ and integrating over the string length yields

$$\left[\mu \int_0^L s_n^2(x) dx \right] \ddot{q}_n(t) + \left[T \int_0^L [s'_n(x)]^2 dx \right] q_n(t) = \int_0^L s_n(x) f_{\text{ext}}(x, t) dx. \quad (4.4)$$

Therefore the equation for the n th mode is that of a second-order oscillator with mass $m_n = \mu \int_0^L s_n^2(x) dx$ and stiffness $k_n = T \int_0^L [s'_n(x)]^2 dx$. The modal shape defines how the external force acts on the mode, and the oscillation $y(x_{\text{out}}, t)$ of the system at a given spatial point x_{out} is the sum of the modal oscillations weighed by the modal shapes: $y(x_{\text{out}}, t) = \sum_{n=1}^{+\infty} s_n(x_{\text{out}}) q_n(t)$.

The two (discrete and continuous) modal representations of oscillating systems have strict analogies, reflecting the fact that continuous systems can be seen as the limit of discrete systems when the number of masses becomes infinite. As an example, a string can be approximated with the discrete network of Fig. 4.5, with N masses. The discrete system has N modes, whose shapes resemble more and more closely those of the continuous system, as N increases.

The modal formalism can be extended to systems that include damping, i.e. where a term $R\dot{y}$ is added in Eq. 4.1, or the terms $d_1 \partial y / \partial t + d_2 \partial / \partial t (\partial^2 y / \partial x^2)$ are added on the left-hand side of Eq. 4.3. However, certain hypotheses about the damping matrix must hold.

Given the modal decomposition for a certain resonating object, sound synthesis can be obtained from a parallel structure of second-order numerical oscillators, each representing a particular mode. Despite the comparatively low computational costs with respect to other techniques, mode-based numerical schemes can become expensive when many objects, each with many modes, are impacted simultaneously. Therefore recent studies deal with optimization of modal synthesis schemes. Bonneel et al. (2008) proposed an approach based on short-time Fourier Transform, that exploits the inherent sparsity of modal sounds in the frequency domain. Other

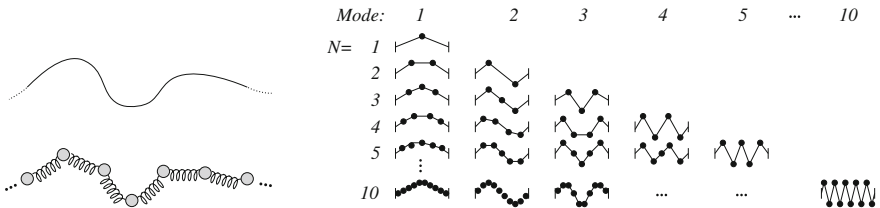


Fig. 4.5 Analogies between continuous and discrete systems. *Left* Approximation of an ideal string with a mass-spring network. *Right* Modes of the discrete system for various numbers N of masses

research has used perceptual criteria to perform mode compression and truncation, so as to reduce the computational load (Raghuvanshi and Lin 2007; Lloyd et al. 2011). The use of variable time-steps in the integration of the numerical equation has also been explored (Zheng and James 2011).

Another current area of research concerns improvements to the basic modal synthesis scheme, with the aim of increasing the realism and quality of the resulting sounds. One of the key challenges is the development of automatic modal analysis and determination of material parameters that recreate realistic audio. Ren et al. (2013) presented a method that analyses pre-recorded target audio clips to estimate perceptually salient modal parameters that capture the inherent quality of recorded sounding materials. A different approach was proposed by Picard et al. (2010), in which automatic voxelization of a surface model is performed, and automatic tuning of the corresponding finite element method parameters is obtained based on the distribution of material in each cell.

4.3.2 Impact Forces

If the external force applied to the resonating object is an ideal impulse, the oscillation is a weighed sum of damped sinusoids. More in general, energy is injected into the system through some kind of excitation mechanism. The amount and the rate at which energy enters the system depends on the nature of the interaction. Impact is a relatively simple interaction, as it occurs in a quasi-impulsive manner, rather than entailing continuous exchange of energy (as it happens for rolling, scraping, stick-slip friction, and so on). At the simplest level, a feed-forward scheme can be used in which the resonator is set into oscillation by driving forces that are externally computed or recorded. As an example, the contact force describing an impact onto a resonating object may be modeled with the following signal (van den Doel and Pai 2004):

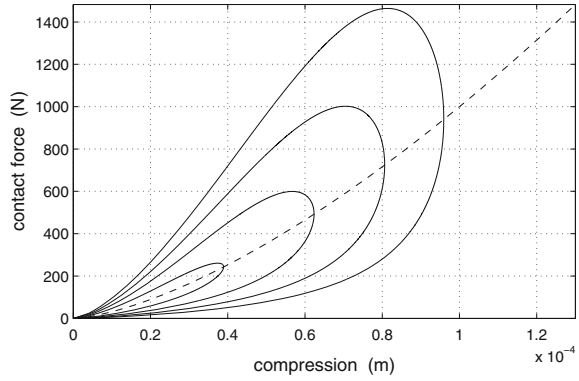
$$f(t) = \begin{cases} \frac{f_{\max}}{2} \left[1 - \cos\left(\frac{2\pi t}{\tau}\right) \right], & 0 \leq t \leq \tau, \\ 0, & t > \tau. \end{cases} \quad (4.5)$$

Here, the time-dependent force signal has a cosinusoidal shape in which the duration of the force (i.e., the contact time) is determined by the parameter τ , while its maximum value is set using the parameter f_{\max} .

As opposed to feed-forward schemes, a more accurate approach to the simulation of contact forces amounts to embedding their computation directly into the model (Avanzini et al. 2003). Despite the complications that arise in the synthesis algorithms, this approach provides some advantages, including improved quality due to accurate audio-rate computation of contact forces, and better interactivity and responsiveness of sound to user actions.

A model for the impact force between two objects, originally proposed by Hunt and Crossley (1975), is the following:

Fig. 4.6 Non-linear force (4.6) generated during impact of a point mass on a hard surface, for various impact velocities



$$f(x(t), v(t)) = \begin{cases} kx(t)^\alpha + \lambda x(t)^\alpha \cdot v(t) & x > 0, \\ 0 & x \leq 0, \end{cases} \quad (4.6)$$

where the time-dependent function $x(t)$ is the interpenetration of the two colliding objects (representing the overall surface deformation of the two objects during contact), and $v = \dot{x}$ is the corresponding velocity. In this case the dynamics of the contact force is not pre-determined as in Eq. 4.5, but is itself dependent on the states and the oscillations of the objects. The parameters k , λ control the stiffness of the impact and the involved dissipation, while the exponent α is related to the local geometry around contact [in particular, $\alpha = 3/2$ in the classic Hertz model of collision between two spheres (Flores et al. 2008)]. Figure 4.6 depicts the simulation of a point mass hitting a rigid surface with the impact model of Eq. 4.6: it can be noted that the impact force has a non-linear characteristics that depends on the exponent α , and exhibits a hysteresis effect that is associated to the dissipative component.

Several refinements to these models have been proposed in order to improve the sound quality. Other relevant phenomena occurring during the impact need to be simulated, particularly the acceleration noise produced as a consequence of large rigid-body accelerations over a short time scale, which results in a perceivable acoustic pressure disturbance at the attack transient (Chadwick et al. 2012).

4.3.3 Rendering of Materials and Hardness in Impact Sounds

As discussed previously, material properties of objects can be perceived auditorily from impact sounds. In particular, object hardness/softness correlates strongly with material identification and impact stiffness.

The modal representation of a resonating object is naturally linked to many ecological dimensions of the corresponding sounds. The possibility of linking the physical model parameters to sound parameters related to the perception of material was first demonstrated by Klatzky et al. (2000). In this work, the modal representation

proposed by van den Doel and Pai (1998) was applied to the synthesis of impact sounds with material information.

The impact force model of Eq. 4.6 was used (Avanzini and Rocchesso 2001; Rocchesso et al. 2003) to produce synthetic auditory stimuli for the investigation of material identification through listening tests. While results from perceptual experiments have already been discussed, here the sound synthesis approach is briefly summarized. In order to minimize the number of model parameters, the modal resonator in the synthesis algorithm was parameterised to have only one mode (i.e., $N = 1$ in Eq. 4.2): as a result, only two acoustic parameters controlled the sound synthesis, namely the centre frequency and the quality factor of the single mode. As a consequence of using a realistic force model, the stimuli maintained the characteristics of impact sounds, despite the “cartoon-like” oversimplification of the resonator.

Ren et al. (2013) investigated the use of the Rayleigh damping model in modal sound synthesis. As discussed in the previous section, modal decoupling is only feasible under certain assumptions for the damping matrix. Rayleigh damping is a commonly adopted approximation model that enables such decoupling, and assumes the damping matrix to be a linear combination of the mass and stiffness matrices. With the goal of investigating whether auditory material perception under the Rayleigh damping assumption is geometry-invariant (i.e., whether this approximation is transferable across different shapes and sizes), Ren et al. (2013) used both real-world audio recordings and (modal) synthesized sounds to perform objective and subjective analysis of the validity of the Rayleigh damping model across different object shapes and sizes. Statistical analysis showed that this is the case for homogeneous materials, whereas the Rayleigh damping model does not provide equally good approximation for materials with heterogeneous micro-structures, such as wood. This study therefore points to some often overlooked limitations of modal sound synthesis.

Aramaki et al. (2011) proposed a modal-based synthesizer of impact sounds, controlled through high-level verbal descriptors referring to material categories (i.e., wood, metal and glass). Analysis was conducted on a set of acoustical descriptors (namely, attack time, spectral bandwidth, roughness, and normalized sound decay), together with electrophysiological measurements (in particular, analysis of changes in brain electrical activity using event related potentials). Based on acoustic and electrophysiological results, a three-layer control architecture providing the synthesis of impact sounds directly from the material label was proposed.

In a study on material perception in a bimodal virtual environment, specifically on the mutual interaction of audio and graphics, Bonneel et al. (2010) also used the modal approach proposed by van den Doel et al. (2001). An experiment similar to the one reported by Klatzky et al. (2000) was conducted. Results showed that the proposed bimodal rendering resulted in better perceived quality for both audio and graphics, and that there was a mutual influence of the two modalities on the perception of material similarity.

Recent research has addressed the issue of auditory rendering of materials from the point of view of walking interactions (Steinicke et al. 2013). Fontana and co-workers (Visell et al. 2009; Marchal et al. 2013) published many results about footstep sound design, including the rendering of floor surface material, as well as multimodal

issues and the integration of vibrotactile display. All sounds were designed using the Sound Design Toolkit (SDT), a software package providing a palette of virtual lutheries and foley pits, targeted at sonic interaction design research and education (Delle Monache et al. 2010). Real-time synthesis of footstep sounds for various materials was also investigated by Nordahl et al. (2010), using modal synthesis and the impact model of Eq. 4.6. A perceptual study was conducted with three groups of subjects: the first group listened to recorded footstep sounds, the second one generated synthetic footstep sounds interactively by walking on shoes augmented with sensors, and the third one listened to pre-recorded footstep sounds synthesized with the same synthesis engine. Results showed that subjects of the second group were able to identify synthesized floor materials at a comparable accuracy with real-world recordings, while the performance of the third group was significantly worse.

Overall, the studies reviewed up to this point show that modal sound synthesis is the most commonly used approach in current literature to render auditorily different materials in impact sounds. There is a trend toward the definition of higher-level control parameters, that refer to ecological categories and hide low-level modal parameters in the background (Aramaki et al. 2011; Delle Monache et al. 2010). Some limitations of this approach have also been highlighted (Ren et al. 2013).

The proportion of studies devoted to the auditory rendering of object hardness/softness is comparatively low with respect to those devoted to material identification. In a modal synthesizer, contact hardness should be rendered by properly adjusting the parameters of the contact force in order to control the hardness-related auditory parameters discussed in the previous sections.

If a physical model such as the one in Eq. 4.6 is used to describe the contact force, an analytical expression for the contact time can be derived (Avanzini and Rocchesso 2004; Papetti et al. 2011):

$$\tau = \left(\frac{m}{k}\right)^{\frac{1}{\alpha+1}} \cdot \left(\frac{\mu^2}{\alpha+1}\right)^{\frac{\alpha}{\alpha+1}} \cdot \int_{v_{out}}^{v_{in}} \frac{1}{(1+\mu v) \left[-\mu(v - v_{in}) + \log \left| \frac{1+\mu v}{1+\mu v_{in}} \right| \right]^{\frac{\alpha}{\alpha+1}}} dv, \quad (4.7)$$

where m is the hammer mass, while the remaining parameters are part of Eq. 4.6. This equation states that the contact time τ depends only on μ , the exponent α and the ratio m/k , in addition to the impact velocity v_{in} . Since neither m nor k affect the value of the integral, it follows that, for a given value of v_{in} , the dependence $\tau \sim (m/k)^{1/(\alpha+1)}$ holds.

Based on this analytical property, a relation between the contact time and the time-varying spectral centroid of the impact sound was discussed by Avanzini and Rocchesso (2004). As a result, a mapping between the physical parameters of the impact force and the hardness-related auditory parameters was proposed. Avanzini and Crosato (2006a) tested this relation in a bimodal (audio-haptic) setting. A subjective test was conducted in which subjects had to tap on virtual audio-haptic surfaces. In each condition the haptic stiffness had the same value while the acoustic stiffness k was varied. Perceived hardness/softness was determined using an absolute

magnitude-estimation procedure. Results showed that subjects consistently ranked the surfaces according to the values of k in the auditory stimuli. If the impact force is not described physically but rather with a signal model, similar control can be achieved. As an example, the cosinusoidal impact force model in Eq. 4.5 includes the contact time τ among its parameters.

As a conclusion to this section, it may be argued that other features of an impact model should be adjusted to properly render the perception of object hardness/softness. As an example, the amount of acceleration noise at the attack (Chadwick et al. 2012) is related to the contact stiffness. Moreover, more complex impact force models (e.g. distributed models taking into account the contact area) may be needed for a more accurate rendering of the attack transient.

4.3.4 Rendering of Deformable and Aggregate Objects

It should be no surprise that the particular topic of rendering sounds produced by interaction with deformable objects, like textiles, tissues, and so on, has seen very little work. In fact, as already noted previously in this chapter, softness-related auditory information is less accessible when interacting with soft objects than with hard ones.

A few studies deal with the topic of textile sounds synthesis, although the relation between the sound rendering and the perceived softness is not investigated. Existing studies have an applicative focus, in which multimodal rendering of textiles is used for high-quality animation and possibly for enhanced active exploration of virtual fabrics (e.g. for e-commerce applications).

Huang et al. (2003) proposed an audio-haptic interface for simulating interaction with a fabric surface through a stylus. The exploratory procedure considered in this work was mainly rubbing of the stylus over a cloth patch. Sound was synthesized using a modal model driven by measured roughness profiles. While appropriate for the particular haptic application considered in this work, the model is hardly generalizable to more complex cloth animations.

It has already been mentioned previously that sliding friction due to textile rubbing against itself or other surfaces is an important component of textile sound but is not the only one. Moreover, frictional sounds are related to surface geometry properties of the object, rather than material properties. In addition to frictional sounds, textiles can also buckle and produce crumpling sounds, in the form of small audible pops. Woven garments produce audible crumpling sounds, while stiff synthetic clothes (e.g. nylon windbreakers), exhibit characteristically loud crumpling sounds. An et al. (2012) proposed a data-driven method for automatically synthesizing sound in physics-based cloth animations. Given a cloth animation, analysis of the deformation was used to drive crumpling and friction sound models estimated from cloth measurements and to synthesize low-quality audio. This was then used as a target signal for a sound synthesis process, which selected best-match short segments from a database of recorded cloth sounds.

Crumpling sounds, such as those used by An et al. (2012) as a component of cloth sound synthesis, are another interesting category of sounds related to softness/hardness perception. An example are the sounds produced by crumpling paper, which can be modeled in terms of (i) the probabilistic distribution of the energies of the short transients, and (ii) a model of the temporal density of transients as a stationary Poisson process (Houle and Sethna 1996). This approach has inspired the development of geometry-independent stochastic models of crumpling, which were used to design sounds produced by deformations of aggregate materials, such as sand, snow, or gravel (Fontana and Bresin 2003). Such sounds belie a common temporal process originating with the transition toward a minimum-energy configuration of an ensemble of microscopic systems, by way of a sequence of transient events. Models of this type have been used in particular to mimic the sound of a footstep onto aggregate grounds (Fontana and Bresin 2003; Marchal et al. 2013; Nordahl et al. 2010; Visell et al. 2009).

This brief overview shows that there is wide space for novel research on the synthesis of sounds of aggregate and deformable materials. The remainder of this section reviews a few studies in which auditory information related to object softness is rendered using different, non-ecological forms of auditory feedback, mostly through some kind of sonification of haptic signals. It may be argued that the adoption of such non-ecological approaches is due to the scarce availability and exploitability of auditory information in the interaction with very soft materials.

Yao et al. (2005) developed a probe to enhance tactile sensations experienced during surgery, specifically during tissue examination with minimally invasive procedures. The probe detects and magnifies the acceleration signal resulting from the interaction of its tool tip with the tissue surfaces. Since the acceleration signal is highly structured and spectrally rich, auditory feedback was obtained through direct conversion of this signal into audio. In the literature of auditory display, this particular approach to sonification is known as “audification” (Dombois and Eckel 2011). Subjective experiments under various conditions (with no amplification, with enhanced tactile feedback, with sound feedback, and with passive touch) showed significant improvements in the recognition of tissue features in the case of tactile and auditory feedbacks.

Kitagawa et al. (2005) performed subjective tests on the sensory substitution of force feedback with sound, in the context of a robotic surgical system. The sound design is not explained in detail: the authors write about a “single tone” (possibly a sinusoid or other waveform) to be played back when the tension applied by the operator exceeded a target value. It was reported that this type of sensory substitution provided statistically significant improvements in applied force accuracy and consistency during the performance of a simple surgical task.

An attempt to formalize a unified approach to study the relationship between physical parameters and coding parameters used to convey control information through the auditory modality was provided by Csapo and Baranyi (2010). The proposed sonification formalism was demonstrated through an application in which the physical properties of a surface are conveyed to a remote teleoperator through sound. Softness/hardness properties, in particular, were continuously sonified using

frequency-modulated or amplitude-modulated signals. However, no subjective tests were conducted to assess the effectiveness of the proposed sonification.

4.3.5 Rendering of Liquid Sounds

Compared to the amount of studies devoted to the sound synthesis and rendering of mechanical interactions between solid objects, those dealing with liquid sounds are a small proportion.

Given the great variety of possible liquid sounds (ranging from stochastic sounds such as that of streaming river, to deterministic ones such as dripping), their synthesis remains a complicated task. Existing research has focused on simulating some of the specific mechanisms responsible for sound generation in liquids, particularly bubble formation. After being formed in a liquid, a bubble emits a decaying sinusoidal sound. If bubble formation occurs close enough to the liquid-air interface, the pitch rises as it approaches the surface. The physical mechanism responsible for these sounds is the pulsation of the bubble volume (Minnaert 1933): any bubble being a small compressible air region surrounded by incompressible fluid, it oscillates like a spring amid a liquid domain.

A few recent studies have dealt with bubble sound synthesis. The first model was proposed in van den Doel's seminal work (van den Doel 2004, 2005). Starting from the physical description provided by Minnaert (1933), a simple algorithm was developed to synthesize single bubbles, using such physical control parameters as bubble radius, loss coefficient, and velocity. Being the model extremely efficient, a real-time bubble simulator was realized, which allowed simulation of more complex liquid sounds (from dripping to heavy rain or waterfalls) through synthesis of a large population of bubbles. The realism of the model was preliminary tested with subjects in a listening experiment. Results suggested that bubbles with radii in the range 2 – 7 mm are most readily associated with the sound of a water drop, and that the rising pitch increases the realism of larger (>4 mm) rather than smaller bubbles, consistently with the fact that these have very high pitch and decay very rapidly. Very large bubbles sounded unnatural, consistently with the fact that they do not occur in isolation in nature.

Zheng and James (2009) proposed a similar approach to acoustic bubble simulation, with the aim of augmenting existing numerical solvers for incompressible liquid simulations that are commonly adopted in the computer graphics literature. The proposed model included bubble advection, time-dependent pitch, and a simplified description of the bubble entrainment process. Sound radiation was modeled through a time-varying sum of bubble oscillators, weighted by their acoustic transfer function modeled as a discrete Green's function of the Helmholtz equation. A fast numerical solver was proposed, which allowed simulation of large numbers of bubbles. Examples for various liquid sounds were proposed (including pouring, babbling, and splashing phenomena), although no psychophysical validation was presented.

Moss et al. (2010) also proposed a simplified, physically inspired model for bubble creation, designed specifically for real-time applications. The model used the fluid surface curvature and velocity as parameters for bubble creation and a stochastic model for bubble sound synthesis based on van den Doel's work (van den Doel 2004, 2005). A user study was conducted to assess the realism of various types of liquid sounds synthesized using the proposed approach. Results suggested that the perceived realism is comparable to recorded sounds in similar settings. However, the model was designed for a shallow water simulator, which reduces interaction possibilities by allowing only surface waves, precluding splashes and object penetration.

In order to bridge the complexity of fluid-dynamic simulations with the needs of interactive sonification, Drioli and Rocchesso (2012) proposed a multi-rate approach to the sound synthesis of liquid phenomena, in which smoothed particle motion simulated at low-rate is used to model liquids in motion and to control audio-rate sound synthesis algorithms of basic acoustic events. Two such basic events were simulated, namely bubbles and surface impacts. In this way, a larger family of sounds can be rendered, including liquid-liquid and liquid-solid interactions. The approach was illustrated through two configurations: the falling of a liquid volume into a container, and the falling of a solid object into a container filled with liquid at rest.

More recently, Cirio et al. (2013) introduced the use of vibrotactile feedback as a rendering modality for solid-fluid interaction, based on the associated sound generating physical processes. Similarly to earlier works, sound was generated from bubble simulation inspired by Moss et al. (2010) and based on a particle-based fluid model (Monaghan 1992). A novel vibrotactile model was then introduced, which received events from the physical simulation and synthesized a signal through three different components: a high-frequency component produced by initial impact of an object onto the liquid, components due to oscillations of smaller bubbles, and the main cavity oscillation. A pilot study was conducted to assess the perceived interaction qualitatively.

As a conclusion to this section, it should be noted that, apart from preliminary user tests aimed at assessing the perceived realism of the proposed simulations, none of the above studies included more extensive psychophysical experiments on the ability of listeners to estimate specific properties of synthesized liquid sound sources.

4.4 Conclusion

Material properties have a front-row seat in the theoretical and empirical study of non-vocal everyday sound sources. Source-perception research has revealed a great deal about the strengths and weaknesses of the auditory estimation of material properties, about how material properties interact perceptually with other mechanical properties of sound-generating events, and about the acoustical factors that underlie perceptual judgments. Most of the research up to this point has focused on stiff solid objects, and has largely disregarded deformable materials such as fabrics or liquids. For both of these, it is thus still unclear the extent to which source-perception

processes might actually rely on material properties rather than material-independent properties such as the texture-defining geometry of fabrics, or the temporally variable geometry of sound-generating bubbles in a liquid. The study of auditory materials has witnessed a number of interesting recent developments that show promising potentials for future research. Research on auditory contexts has begun to unravel the factors involved in the more general ability to differentiate between states of matter of sound-generating substances, and to benchmark the perception of materials against that of the properties of sound-generating interactions. Research on audio-haptic contexts has begun to address the interactions between material information presented in different modalities from both the perceptual and motor-control points of view. Further, promising directions of research are also represented by the study of the cortical processes involved in the processing of material-related information (Arnott et al. 2008; Aramaki et al. 2010; Micoulaud-Franchi et al. 2011).

Studies in ecological acoustic have been re-discovered in the late 1990's in the light of sound design and sound rendering for virtual reality, and have been a major driver for research on the synthesis non-vocal everyday sounds. Techniques for modal synthesis of sounds produced by stiff objects in impulsive or continuous contact are now well established. It can be expected that upcoming research will continue to focus on the development of more refined and realistic models of the interaction. Due to the impulsive and highly non-linear nature of impact forces, one current open issue concerns the definition of specialized numerical techniques for the accurate simulation of such forces (Papetti et al. 2011; Chatziioannou and van Walstijn 2013). Further improvements in realism will be achieved through the simulation of secondary physical mechanisms involved in the interaction, such as acceleration noise and its relation to the contact stiffness (Chadwick et al. 2012), as well as the effects of distributed and possibly time-varying contact areas.

In the mid to long term, it can be expected that other physical modeling techniques, such as time-domain finite differences (FDTD) and finite element methods (FEM) will gain popularity and become competitive with modal synthesis. Being "brute force" approaches, they possess the advantage of generality since a great variety of systems can be approached without the need for making simplifying hypotheses or adding intermediate levels of representation. Material properties in particular are completely controllable, since all low-level material-related parameters (Young and shear modulus, Poisson coefficient, density, etc.) are directly embedded into the models. On the other hand, such methods are numerically intensive: future research will therefore be devoted to looking at efficient implementations particularly in parallel architectures (multicore processors and general purpose graphics processing units, see e.g. Bilbao et al. 2013). Improvements in sound quality promise to be striking, however only very recently has computational power grown to the extent that sound can be synthesized in a reasonable amount of time with these techniques, and real-time is still a long way off.

While impacts between stiff objects have been thoroughly studied, there is wide space for novel research on the synthesis of other categories of sounds. Our review has shown that there is a handful of research on the synthesis of sounds produced by deformable objects. In particular, the most recent studies on textile and cloth

sounds illustrate the many issues involved in the synthesis of such complex sounds, and at the same time demonstrate the potential for research in this direction. Similar considerations apply to liquid sound synthesis, although in this case the number of existing studies is marginally larger. In both cases, there is a lack of validation of the proposed approaches in terms of their ability to convey specific sound source properties to the listener.

Acknowledgments This work was partially supported by the Marie Curie Intra-European Fellowships program (FP7 PEOPLE-2011-IEF-30153, project BrainInNaturalSound to Bruno L. Gior-dano). The authors wish to thank Laurie Heller and Guillaume Lemaître for sharing the sound stimuli used to prepare Figs. 4.1 and 4.3, and Laurie Heller, Federico Fontana and Stephen McAdams for providing helpful feedback about earlier versions of this chapter.

References

- Adrien J-M (1991) The missing link: modal synthesis. In: De Poli G, Piccialli A, Roads C (eds) *Representations of musical signals*. MIT Press, Cambridge, pp 269–297
- An SS, James DL, Marschner S (2012) Motion-driven concatenative synthesis of cloth sounds. *ACM Trans Graph (TOG)* 31(4) (Article No. 102)
- Aramaki M, Marie C, Kronland-Martinet R, Ystad S, Besson M (2010) Sound categorization and conceptual priming for nonlinguistic and linguistic sounds. *J Cognit Neurosci* 22:2555–2569
- Aramaki M, Besson M, Kronland-Martinet R, Ystad S (2011) Controlling the perceived material in an impact sound synthesizer. *IEEE Trans Audio Speech Lang Process* 19(2):301–314
- Arnott SR, Cant JS, Dutton GN, Goodale MA (2008) Crinkling and crumpling: an auditory fmri study of material properties. *Neuroimage* 43:368–378
- Ashby FG (1992) *Multidimensional models of perception and cognition*. Lawrence Erlbaum Associates, Hills-dale
- Attias H, Schreiner CE (1997) Temporal low-order statistics of natural sounds. In: Mozer MC, Jordan MI, Petsche T (eds) *Advances in neural information processing systems 9*. MIT Press, Cambridge, pp 27–33
- Avanzini F, Rath M, Rocchesso D, Ottaviani L (2003) Low-level sound models: resonators, interactions, surface textures. In: Rocchesso D, Fontana F (eds) *The sounding object*. Mondo Estremo, Firenze, pp 137–172
- Avanzini F, Crosato P (2006b) Integrating physically-based sound models in a multimodal rendering architecture. *Comput Anim Virtual Worlds* 17(3–4):411–419. doi:[10.1002/cav.v17:3/4](https://doi.org/10.1002/cav.v17:3/4)
- Avanzini F, Crosato P (2006a) Haptic-auditory rendering and perception of contact stiffness. In: McGookin D, Brewster S (eds) *Haptic and audio interaction design*. Lecture Notes in Computer Science 4129/2006, Springer, Berlin. pp 24–35
- Avanzini F, Rocchesso D (2001) Controlling material properties in physical models of sounding objects. In: *Proceedings of the international computer music conference (ICMC'01)*. La Habana, pp 91–94 (Available at <http://www.soundobject.org>)
- Avanzini F, Rocchesso D (2004) Physical modeling of impacts: theory and experiments on contact time and spectral centroid. In: *Proceedings of the sound and music computing conference (SMC2004)*, Paris, pp 287–293
- Ballas JA (1993) Common factors in the identification of an assortment of brief everyday sounds. *J Exp Psychol Hum Percept Perform* 19:250–267
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. *Nature* 403:309–312

- Bilbao S (2009) Numerical sound synthesis—finite difference schemes and simulation in musical acoustics. Wiley, Chichester
- Bilbao S, Hamilton B, Torin A, Webb C, Graham P, Gray A, Perry J (2013) Large scale physical modeling sound synthesis. In: Proceedings of the Stockholm music acoustic conference (SMAC2013), Stockholm, pp 593–600
- Bonneel N, Drettakis G, Tsingos N, Viaud-Delmon I, James D (2008) Fast modal sounds with scalable frequency-domain synthesis. *ACM Trans Graph (TOG)* 27(3) (Article no. 24.)
- Bonneel N, Suied C, Viaud-Delmon I, Drettakis G (2010) Bimodal perception of audio-visual material properties for virtual environments. *ACM Trans Appl Percept* 7(1) (Article No. 1)
- Borg I, Groenen P (1997) Modern multidimensional scaling. Springer, New York
- Cabe PA, Pittenger JB (2000) Human sensitivity to acoustic information from vessel filling. *J Exp Psychol Hum Percept Perform* 26:313–324
- Carello C, Wagman JB, Turvey MT (2003) Acoustical specification of object properties. In: Anderson J, Anderson B (eds) Moving image theory: ecological considerations. Southern Illinois University Press, Carbondale
- Castiello U, Giordano BL, Begliomini C, Ansuini C, Grassi M (2010) When ears drive hands: the influence of contact sound on reaching to grasp. *PLoS ONE* 5:e12240
- Chadwick JN, Zheng C, James DL (2012) Precomputed acceleration noise for improved rigid-body sound. *ACM Trans Graph* 31(4) (Article No. 103)
- Chatzioannou V, van Walstijn M (2013) An energy conserving finite difference scheme for simulation of collisions. In: Proceedings of the sound and music computing conference (SMC2013), Stockholm, pp 584–591
- Cho G, Casali JG, Yi E (2001) December). Effect of fabric sound and touch on human subjective sensation. *Fibers Polym* 2(4):196–202
- Cho G, Kim C, Cho J, Ha J (2005) March). Physiological signal analyses of frictional sound by structural parameters of warp knitted fabrics. *Fibers Polym* 6(1):89–94
- Cirio G, Marchal M, Lécuyer A, Cooperstock J (2013) Vibrotactile rendering of splashing fluids. *ACM Trans Haptics* 6(1):117–122
- Civille GV, Dus CA (1990) Development of terminology to describe the handfeel properties of paper and fabrics. *J Sen Stud* 5(1):19–32
- Csapo AB, Baranyi P (2010) An interaction-based model for auditory substitution of tactile percepts. In: Proceedings of IEEE international conference on intelligent engineering systems (INES 2010), Las Palmas of Gran Canaria, pp 248–253
- De Coensel B, Vanwetswinkel S, Botteldooren D (2011) Effects of natural sounds on the perception of road traffic noise. *J Acoust Soc Am* 129:EL148–EL153
- Delle Monache, S., Polotti, P., & Rocchesso, D. (2010, September). A toolkit for explorations in sonic interaction design. In: Proceedings of audio mostly conference (AM'10). Piteå (Article no. 1)
- DiFranco DE, Beauregard GL, Srinivasan MA (1997) The effect of auditory cues on the haptic perception of stiffness in virtual environments. In: Proceedings of the ASME dynamic systems and control division, DSC, vol 61, pp 17–22
- Dombois F, Eckel G (2011) Audification. In: Hermann T, Hunt A, Neuhoff JG (eds) The sonification handbook. Logos Verlag, Berlin, pp 301–324
- Doutaut V, Matignon D, Chaigne A (1998) Numerical simulations of xylophones. II. Time-domain modeling of the resonator and of the radiated sound pressure. *J Acoust Soc Am* 104(3):1633–1647
- Drioli C, Rocchesso D (2012) Acoustic rendering of particle-based simulation of liquids in motion. *J Multimodal User Interfaces* 5(3–4):187–195
- Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415:429–433
- Fletcher NH, Rossing TD (1991) The physics of musical instruments. Springer, New York
- Flores P, Claro JP, Lankarani HM (2008) Kinematics and dynamics of multibody systems with imperfect joints: Models and case studies. Springer, Berlin

- Fontana F, Bresin R (2003) Physics-based sound synthesis and control: crushing, walking and running by crumpling sounds. In: *Proceedings of the colloquium on music informatics (CIM 2003)*, Firenze, pp 109–114
- Franić K, Serafin S (eds) (2013) *Sonic interaction design*. MIT Press, Cambridge
- Freed DJ (1990) Auditory correlates of perceived mallet hardness for a set of recorded percussive events. *J Acoust Soc Am* 87:311–322
- Gaver WW (1988) *Everyday listening and auditory icons*. Unpublished doctoral dissertation, University of California, San Diego
- Gaver WW (1993) What in the world do we hear? an ecological approach to auditory event perception. *Ecol Psychol* 5:1–29
- Geffen MN, Gervain J, Werker JF, Magnasco MO (2011) Auditory perception of self-similarity in water sounds. *Front Integr Neurosci* 5
- Gibson JJ (1966) *The senses considered as a perceptual system*. Houghton Mifflin, Boston
- Gibson JJ (1979) *The ecological approach to visual perception*. Houghton Mifflin, Boston
- Giordano BL, Avanzini F, Wanderley M, McAdams S (2010) Multisensory integration in percussion performance. In: *Actes du 10eme congrès français d'acoustique*, Lyon (p. [CD-ROM]). Société Française d'Acoustique, Paris, France
- Giordano BL (2005) *Sound source perception in impact sounds*. Unpublished doctoral dissertation, University of Padova, Italy
- Giordano BL (2003) Material categorization and hardness scaling in real and synthetic impact sounds. In: Rocchesso D, Fontana F (eds) *The sounding object*. Mondo Estremo, Firenze, pp 73–93
- Giordano BL, McAdams S (2006) Material identification of real impact sounds: effects of size variation in steel, glass, wood and plexiglass plates. *J Acoust Soc Am* 119:1171–1181
- Giordano BL, McDonnell J, McAdams S (2010) Hearing living symbols and nonliving icons: category-specificities in the cognitive processing of environmental sounds. *Brain Cogn* 73:7–19
- Giordano BL, Rocchesso D, McAdams S (2010) Integration of acoustical information in the perception of impacted sound sources: the role of information accuracy and exploitability. *J Exp Psychol Hum Percept Perform* 36:462–479
- Giordano BL, Guastavino C, Murphy E, Ogg M, Smith BK, McAdams S (2011) Comparison of methods for collecting and modeling dissimilarity data: applications to complex sound stimuli. *Multivar Behav Res* 46:779–811
- Giordano BL, Visell Y, Yao HY, Hayward V, Cooperstock J, McAdams S (2012) Identification of walked-upon materials in auditory, kinesthetic, haptic and audio-haptic conditions. *J Acoust Soc Am* 131:4002–4012
- Grassi M (2005) Do we hear size or sound? Balls dropped on plates. *Percept Psychophys* 67:274–284
- Gygi B, Kidd GR, Watson CS (2007) Similarity and categorization of environmental sounds. *Percept Psychophys* 69:839–855
- Handel S, Erickson ML (2001) A rule of thumb: the bandwidth for timbre invariance is one octave. *Music Percept* 19:121–126
- Houix O (2003) *Categorisation auditive des sources sonores*. Unpublished doctoral dissertation, Université du Maine, France
- Houix H, Lemaitre G, Misdariis N, Susini P, Urdapilleta I (2012) A lexical analysis of environmental sound categories. *J Exp Psychol Appl* 18:52
- Houle PA, Sethna JP (1996) Acoustic emission from crumpling paper. *Phys Rev E* 54(1):278–283
- Huang G, Metaxas D, Govindaraj M (2003) Feel the "fabric": an audio-haptic interface. In: *Proceedings of ACM siggraph/eurographics symposium on computer animation (SCA03)*. San Diego, pp 52–61
- Hunt KH, Crossley FRE (1975) Coefficient of restitution interpreted as damping in vibroimpact. *J Appl Mech* 42:440–445
- Jansson G (1993) Perception of the amount of fluid in a vessel shaken by hand. In: Valenti S, Pittinger J (eds) *Studies in perception and action II*. Posters presented at the VIIth international conference on event perception and action, pp 263–267

- Jansson G, Juslin P, Poom L (2006) Liquid-specific stimulus properties can be used for haptic perception of the amount of liquid in a vessel put in motion. *Perception* 35:1421–1432
- Jeon JY, Lee JL, You J, Kang J (2012) Acoustical characteristics of water sounds for soundscape enhancement in urban open spaces. *J Acoust Soc Am* 131:2101–2109
- Kawabata S (1980) The standardization and analysis of hand evaluation. Textile Machinery Society of Japan, Osaka
- Kidd GR, Watson CS (2003) The perceptual dimensionality of environmental sounds. *Noise Control Eng J* 51:216–231
- Kitagawa M, Dokko D, Okamura AM, Yuh DD (2005) Effect of sensory substitution on suture-manipulation forces for robotic surgical systems. *J Thorac Cardiovasc Surg* 129(1):151–158
- Klatzky RL, Pai DK, Krotkov EP (2000) Perception of material from contact sounds. *Presence Teleoperators Virtual Environ* 9(4):399–410
- Kunkler-Peck AJ, Turvey MT (2000) Hearing shape. *J Exp Psychol Hum Percept Perform* 26:279–294
- Lakatos S, McAdams S, Caussé R (1997) The representation of auditory source characteristics: simple geometric form. *Percept Psychophys* 59:1180–1190
- Lambourg C, Chaigne A, Matignon D (2001) Time-domain simulation of damped impacted plates: II. Numerical Models and Results. *J Acoust Soc Am* 109(4):1433–1447
- Lederman SJ (1979) Auditory texture perception. *Perception* 8(1):93–103
- Lederman SJ, Klatzky RL (2004) Multisensory texture perception. In: Calvert G, Spence C, Stein B (eds) *Handbook of multisensory processes*. MIT Press, Cambridge, pp 107–122
- Lemaitre G, Heller LM (2012) Auditory perception of material is fragile while action is strikingly robust. *J Acoust Soc Am* 131:1337–1348
- Lemaitre G, Heller LM (2013) Evidence for a basic level in a taxonomy of everyday action sounds. *Exp Brain Res* 226(2):253–264
- Lewis JW, Brefczynski JA, Phinney RE, Jannik JJ, DeYoe ED (2005) Distinct cortical pathways for processing tool versus animal sounds. *J Neurosci* 25:5148–5158
- Lloyd B, Raghuvanshi N, Govindaraju NK (2011) Sound synthesis for impact sounds in video games. In: *Proceedings of the ACM symposium on interactive 3D graphics and games (ACM I3D)*, San Francisco, pp 55–62
- Lutfi RA, Oh EL (1997) Auditory discrimination of material changes in a struck-clamped bar. *J Acoust Soc Am* 102:3647–3656
- Lutfi RA (2001) Auditory detection of hollowness. *J Acoust Soc Am* 110:1010–1019
- Lutfi RA (2007) Human sound source identification. In: Yost WA, Fay RR, Popper AN (eds) *Auditory Percept Sound Sources*. Springer, New York, NY, pp 13–42
- Lutfi RA, Liu CJ (2007) Individual differences in source identification from synthesized impact sounds. *J Acoust Soc Am* 122:1017–1028
- Lutfi RA, Stoelinga CNJ (2010) Sensory constraints on auditory identification of the material and geometric properties of struck bars. *J Acoust Soc Am* 127:350–360
- Marchal M, Cirio G, Visell Y, Fontana F, Serafin S, Cooperstock J, Lécuyer A (2013) Multimodal rendering of walking over virtual grounds. In: Steinicke F, Visell Y, Campos J, Lécuyer A (eds) *Human walking in virtual environments*. Springer, New York, pp 263–295
- Marozeau J, de Cheveigné A, McAdams S, Winsberg S (2003) The dependency of timbre on fundamental frequency. *J Acoust Soc Am* 114:2946–2957
- McAdams S, Chaigne A, Roussarie V (2004) The psychomechanics of simulated sound sources: material properties of impacted bars. *J Acoust Soc Am* 115:1306–1320
- McAdams S, Roussarie V, Chaigne A, Giordano BL (2010) The psychomechanics of simulated sound sources: material properties of impacted thin plates. *J Acoust Soc Am* 128:1401–1413
- Michaels CF, Carello C (1981) *Direct perception*. Prentice-Hall, Englewood Cliffs
- Micoulaud-Franchi J-A, Aramaki M, Merer A, Cermolacce M, Ystad S, Kronland-Martinet R, Vion-Dury J (2011) Categorization and timbre perception of environmental sounds in schizophrenia. *Psychiatry Res* 189:149–152

- Minnaert M (1933) On musical air-bubbles and the sounds of running water. *Lond Edinb Dublin Philos Maga J Sci* 16:235–248
- Monaghan JJ (1992) Smoothed particle hydrodynamics. *Ann Rev Astron Astrophys* 30:543–574
- Moss W, Yeh H, Hong J-M, Lin MC, Manocha D (2010) Sounding liquids: automatic sound synthesis from fluid simulation. *ACM Trans Graph* 29(3) (Article No. 21)
- Nordahl R, Serafin S, Turchet L (2010) Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In: *Proceedings of the IEEE international conference on virtual reality (VR 2010)*, Waltham, pp 147–153
- O'Brien JF, Shen C, Gatchalian CM (2002) Synthesizing sounds from rigid-body simulations. In: *Proceedings of the 2002 ACM SIGGRAPH/Eurographics symposium on computer animation*. San Antonio, TX, pp 175–181
- Papetti S, Avanzini F, Rocchesso D (2011) Numerical methods for a non-linear impact model: a comparative study with closed-form corrections. *IEEE Trans Audio Speech Lang Process* 19(7):2146–2158
- Picard C, Frisson C, Faure F, Drettakis G, Kry P (2010) Advances in modal analysis using a robust and multiscale method. *EURASIP J Adv Sig Process* 2010 (Article ID 392782.)
- Raghuvanshi N, Lin MC (2007) Physically based sound synthesis for large-scale virtual environments. *Comput Graph Appl IEEE* 27(1):14–18
- Ren Z, Yeh H, Klatzky R, Lin MC (2013) Auditory perception of geometry-invariant material properties. *IEEE Trans Vis Comput Graph* 19(4):557–566
- Ren Z, Yeh H, Lin MC (2013) Example-guided physically based modal sound synthesis. *ACM Trans Graph (TOG)* 32(1) (Article No. 1)
- Rocchesso D, Ottaviani L, Fontana F, Avanzini F (2003) Size, shape, and material properties of sound models. In: Rocchesso D, Fontana F (eds) *The sounding object*. Mondo Estremo, Firenze, pp 95–110
- Rocchesso D (2011) *Explorations in sonic interaction design*. Logos Verlag, Berlin
- Smith JO III (2004) Virtual acoustic musical instruments: review and update. *J New Music Res* 33(3):283–304
- Sreng J, Bergez F, Legarrec J, Lécuyer A, Andriot C (2007) Using an event-based approach to improve the multimodal rendering of 6dof virtual contact. In: *Proceedings of the ACM symposium on virtual reality software and technology (VRST07)*, Newport Beach, CA, pp 165–173
- Steele KM, Williams AK (2006) Is the bandwidth for timbre invariance only one octave? *Music Percept* 23:215–220
- Steinicke F, Visell Y, Campos J, Lecuyer A (eds) (2013) *Human walking in virtual environments: perception, technology, and applications*. Springer, New York
- Tucker S, Brown GJ (2003) Modelling the auditory perception of size, shape and material: applications to the classification of transient sonar sounds. In: *Proceedings of the 114th convention of the Audio Engineering Society*
- Turchet L, Serafin S (2014) Semantic congruence in audio-haptic simulation of footsteps. *Appl Acoust* 75:59–66
- Turchet L, Serafin S, Cesari P (2013) Walking pace affected by interactive sounds simulating stepping on different terrains. *ACM Trans Appl Percept* 10:23
- Välimäki V, Pakarinen J, Erku C, Karjalainen M (2006) Discrete-time modelling of musical instruments. *Reports Prog Phys* 69(1):1–78
- van den Doel K (2004) Physically-based models for liquid sounds. In: *Proceedings of the international conference on auditory display (ICAD04)*, Sydney
- van den Doel K (2005) Physically based models for liquid sounds. *ACM Trans Appl Percept* 2:534–546
- van den Doel K, Pai DK (1998) The sounds of physical shapes. *Presence Teleoperators Virtual Environ* 7(4):382–395
- van den Doel K, Pai DK (2004) Modal synthesis for vibrating objects. In: *Greenebaum K (ed) Audio anecdotes*. AK Peters, Natick

- van den Doel K, Kry PG, Pai DK (2001) FoleyAutomatic: physically-based sound effects for interactive simulation and animation. In: Proceedings of the international conference on computer graphics and interactive techniques (SIGGRAPH 01), Los Angeles, CA, pp 537–544
- Vanderveer NJ (1979) Ecological acoustics: Human perception of environmental sounds. Unpublished doctoral dissertation, Cornell University. (Dissertation Abstracts International, 40, 4543B. (University Microfilms No. 80–04–002))
- Velasco C, Jones R, King S, Spence C (2013) The sound of temperature: What information do pouring sounds convey concerning the temperature of a beverage. *J Sens Stud* 28:335–345
- Visell Y, Fontana F, Giordano BL, Nordahl R, Serafin S, Bresin R (2009) Sound design and perception in walking interactions. *Int J Hum Comput Stud* 67(11):947–959
- Voss RF, Clarke J (1975) 1/f noise in speech and music. *Nature* 258:317–318
- Warren WH, Verbrugge RR (1984) Auditory perception of breaking and bouncing events: a case study in ecological acoustics. *J Exp Psychol Hum Percept Perform* 10:704–712
- Waterman NA, Ashby MF (1997) The materials selector, 2nd edn. Chapman and Hall, London
- Wildes R, Richards W (1988) Recovering material properties from sound. In: Richards W (ed) *Nat Comput*. MIT Press, Cambridge, MA, pp 356–363
- Yao H-Y, Hayward V, Ellis RE (2005) A tactile enhancement instrument for minimally invasive surgery. *Comput Aided Surg* 10(4):233–239
- Zheng C, James DL (2009) Harmonic fluids. *ACM Trans Graph (TOG)* 28(3) (Article No. 37)
- Zheng C, James DL (2011) Toward high-quality modal contact sound. *ACM Trans Graph (TOG)* 30(4) (Article No. 38)